

A data-driven test for cross-cultural differences in face preferences

Lingshan Zhang¹, Iris J Holzleitner¹, Anthony J Lee², Hongyi Wang³,
Chengyang Han¹, Vanessa Fasolt¹, Lisa M DeBruine¹, Benedict C Jones¹

1. Institute of Neuroscience & Psychology, University of Glasgow, Scotland, UK

2. Division of Psychology, University of Stirling, Scotland, UK.

3. School of Psychology and Cognitive Science, East China Normal University, China

Data, analysis code, and preregistration: <https://osf.io/7wy3t/>

Email of corresponding author: ben.jones@glasgow.ac.uk

Acknowledgements: This research was funded by ERC grant KINSHIP awarded to LMD.

Zhang L, Holzleitner IJ, Lee AJ, Wang H, Han C, Fasolt V, DeBruine LM & Jones BC, A data-driven test for cross-cultural differences in face preferences, *Perception*, 48 (6), pp. 487-499. Copyright © The Authors 2019. Reprinted by permission of SAGE Publications. DOI: <https://doi.org/10.1177/0301006619849382>

A data-driven test for cross-cultural differences in face preferences

Abstract

Previous research has shown strong cross-cultural agreement in facial attractiveness judgments. However, these studies all used a theory-driven approach in which responses to specific facial characteristics are compared between cultures. This approach is constrained by the predictions that can be derived from existing theories and can therefore bias impressions of the extent of cross-cultural agreement in face preferences. We directly addressed this problem by using a data-driven, rather than theory-driven, approach to compare facial attractiveness judgments made by Chinese-born participants who were resident in China, Chinese-born participants currently resident in the UK, and UK-born and -resident White participants. Analyses of the principal components along which faces naturally varied suggested that Chinese and White UK participants used face information in different ways, at least when judging women's facial attractiveness. In other words, the data-driven approach used in the current study revealed some cross-cultural differences in face preferences that were not apparent in studies using theory-driven approaches.

Introduction

Facial attractiveness judgments influence important social outcomes, including hiring decisions and interpersonal relationships (Langlois et al., 2000; Little et al., 2011; Rhodes, 2006). Cross-cultural agreement in facial attractiveness judgments is widely interpreted as strong evidence that face preferences transcend culture (Langlois et al., 2000; Rhodes, 2006).

Previous research investigating cross-cultural agreement in facial attractiveness judgments has used a top-down, theory-driven approach (Apicella et al., 2007; Little et al., 2007; Perrett et al., 1998; Rhodes et al., 2000). In this approach, specific characteristics identified from evolutionary theories of attractiveness (e.g., symmetry, averageness, sexual dimorphism, Little et al., 2011 and Thornhill & Gangestad, 1999) are experimentally

manipulated in face images using computer graphics (Apicella et al., 2007; Little et al., 2007; Perrett et al., 1998; Rhodes et al., 2001).

Studies using this theory-driven approach have found that Japanese and Hadza participants showed preferences for facial symmetry and averageness similar to those reported for Western cultures (Apicella et al., 2007; Little et al., 2007; Rhodes et al., 2001). Other studies using this approach found that manipulating sexually dimorphic shape characteristics in face images had similar effects on Japanese and Western participants' attractiveness judgments (Perrett et al., 1998). For example, both Japanese and Western participants preferred feminized versions of faces to masculinized versions (Perrett et al., 1998).

Results like those described above are typically interpreted as evidence for cross-cultural agreement in face preferences (Apicella et al., 2007; Little et al., 2007; Perrett et al., 1998; Rhodes et al., 2001). However, manipulating characteristics such as sexual dimorphism in two-dimensional face images can also alter perceptions of more changeable characteristics, such as head orientation or tilt (see, e.g., Hehman et al., 2013 and Schneider et al., 2012). Moreover, the theory-driven approach used in these studies has two important limitations.

First, the facial characteristics investigated in these studies may not necessarily contribute substantially to facial attractiveness judgments. For example, Said and Todorov (2011) found that the combined effects of sexual dimorphism and averageness explained only ~5% of the variance in women's attractiveness ratings of male face images (see also Holzleitner et al., 2018).

Second, the range of hypotheses that can be tested using the theory-driven approach is constrained by existing theoretical frameworks. Because the ability to detect cultural differences will then depend entirely on which specific stimulus characteristics are manipulated, this constraint can bias our impressions of the extent of cross-cultural agreement in responses to social signals (Jack et al., 2018). By contrast, bottom-up, data-driven approaches do

not have this constraint, meaning that they can reveal cultural differences that existing theories of social perception do not predict (Jack et al., 2018). Indeed, data-driven approaches to studying facial expressions of emotion have revealed cultural differences in emotion perception that were not evident (or predicted) in studies using theory-driven approaches (Jack et al., 2018).

In light of the above, we first used a data-driven approach (Principal Component Analysis, PCA) to identify the principal components (shape PCs) along which face images naturally varied. We then tested whether these PCs predicted Chinese and White UK participants' attractiveness ratings of the faces in different ways. We used attractiveness ratings made by three different groups of participants (White UK-born UK-resident participants, Chinese-born UK-resident participants, and Chinese-born China-resident participants). We tested both Chinese and White UK face images. Methods were preregistered on the Open Science Framework prior to data collection (<https://osf.io/7wy3t/>).

Methods

Face stimuli

Stimuli were face photographs of 50 Chinese men (mean age=24.39 years, SD=3.52 years), 50 Chinese women (mean age=23.94 years, SD=2.63 years), 50 White UK men (mean age=22.97 years, SD=5.95 years), and 50 White UK women (mean age=21.95 years, SD=3.60 years). These men and women first cleaned their face with hypoallergenic face wipes to remove any make-up. Face photographs were taken a minimum of 15 minutes later in a small windowless room against a constant background, and under standardized diffuse lighting conditions. The men and women were instructed to pose with a neutral expression. Camera-to-head distance and camera settings were held constant. Six photographs of each individual were taken simultaneously from a variety of angles. Images were collected using a DI3D system (www.di4d.com) using six standard digital cameras (Canon EOS100D with Canon EF 50 mm f/1.8 STM lenses). Only the front-view face images were used in this study. In this image capture system, camera height is

adjusted for each participant to minimize variation in head tilt due to camera-height.

Face ratings

Faces were rated for attractiveness using a 1 (very unattractive) to 7 (very attractive) scale by 15 Chinese China-resident men (mean age=23.7 years, SD=1.9 years), 15 Chinese China-resident women (mean age=21.7 years, SD=2 years), 15 Chinese UK-resident men (mean age=24.6 years, SD=2.7 years; mean time resident in UK=352 days, SD=652 days), 15 Chinese UK-resident women (mean age=23.8 years, SD=2.7 years; mean time resident in UK=420 days, SD=606 days), 15 White UK men (mean age=21.4 years, SD=2.2 years), and 15 White UK women (mean age=21.4 years, SD=3.5 years). Following previous work that used similar data-driven methods to study Western participants' attractiveness judgments (Said & Todorov, 2011), participants rated the attractiveness of opposite-sex faces only. Trial order was fully randomized. Simulations (see <https://osf.io/x7fus/>) sampling from a population of 2513 raters, each of whom had rated the attractiveness of 102 faces, indicated that >99% of 1000 random samples of 15 raters produced Cronbach's alphas >.8 (90% of all alphas were >.85). This indicates that 15 raters per group are typically sufficient to obtain reliable average ratings. For ratings, each image was standardized on pupil positions and masked so that hairstyle and clothing were not visible.

Consistent with the results of our simulations, inter-rater agreement (Cronbach's alphas) for ratings of individual faces was high for each of the six groups of raters (Chinese China-resident raters judging men's faces=.88; Chinese UK-resident raters judging men's faces=.85; White UK raters judging men's faces=.87; Chinese China-resident raters judging women's faces=.80; Chinese UK-resident raters judging women's faces=.87; White UK raters judging women's faces=.85). For each face, the mean attractiveness rating was calculated separately from each group's ratings (Chinese China-resident raters, Chinese UK-resident raters, White UK raters). These mean ratings served as the dependent variables in our analyses. Following previous research that used similar data-driven methods to study Western participants'

attractiveness judgments (Holzleitner et al., 2018; Said & Todorov, 2011), raw ratings were standardized (converted to z scores) prior to averaging. Before standardizing, ratings were similar to those reported for attractiveness in studies using similar stimuli (Bronstad et al., 2008; Kościński, 2013; Torrance et al., 2014; Wang et al., 2016; see Table 1).

Table 1. Descriptive statistics for attractiveness ratings. Table shows means (and standard deviation in parentheses). Descriptive statistics are for raw ratings.

	White UK raters	Chinese UK-resident raters	Chinese Chinese-resident raters
Chinese male faces	2.82 (0.57)	2.32 (0.48)	2.54 (0.50)
Chinese female faces	2.85 (0.61)	2.95 (0.69)	2.80 (0.51)
White UK male faces	3.12 (0.74)	2.81 (0.56)	3.24 (0.64)
White UK female faces	2.80 (0.66)	3.05 (0.51)	2.83 (0.46)

Principal Component Analysis (PCA) of faces

Orthogonal face principal components (PCs) were derived from 132 points on each of the 200 faces using a method described in Wolffhechel et al. (2015). Note that this is a larger number of images than has been used to derive face PCs in many previous studies (e.g., Holzleitner et al., 2014; Komori et al., 2011; Scott et al., 2010). Images were Procrustes aligned prior to analyses (using the 2D images, following, e.g., Scott et al., 2010). The image-analysis code used to calculate these face PCs is publicly available on the Open Science Framework (<https://osf.io/7wy3t/>).

Table 2. Average eigenvalues for first three PCs by face group.

face ethnicity	face sex	PC1	PC2	PC3
----------------	----------	-----	-----	-----

Chinese	Female	0.49583770	0.7619386	0.13630340
Chinese	Males	-0.08793491	0.5106452	-0.19655207
White UK	Female	-0.01282015	-0.2182166	-0.05068316
White UK	Male	-0.39508264	-1.0543672	0.11093182

We used the broken stick criterion to select the PCs to be included as predictors in our preregistered analyses (see Jackson, 1993 for a discussion of the benefits of this criterion). This method selected 12 PCs, cumulatively explaining 81% of the variance in 2D face shape. The first three of these PCs, which explained 48% of the variance in 2D face shape (27%, 11%, and 10%, respectively), are visualized in Figure 1 (visualizations of all 12 PCs are at <https://osf.io/7wy3t/>). These three PCs appeared to reflect head tilt and sexual dimorphism, face ethnicity, and elongation, respectively. PCs 1 and 3 are similar to those reported in previous work on PCAs of White faces (e.g., Hancock et al., 1998). PC2 is presumably a consequence of including two distinct racial groups in our image set. Average eigenvalues for PCs 1 to 3 for each of the four face groups are shown in Table 2 (shown for all 12 shape PCs at <https://osf.io/7wy3t/>). Conducting the PCA on male and female face shapes separately revealed similar PCs 1 to 3 (see <https://osf.io/7wy3t/> for visualizations).

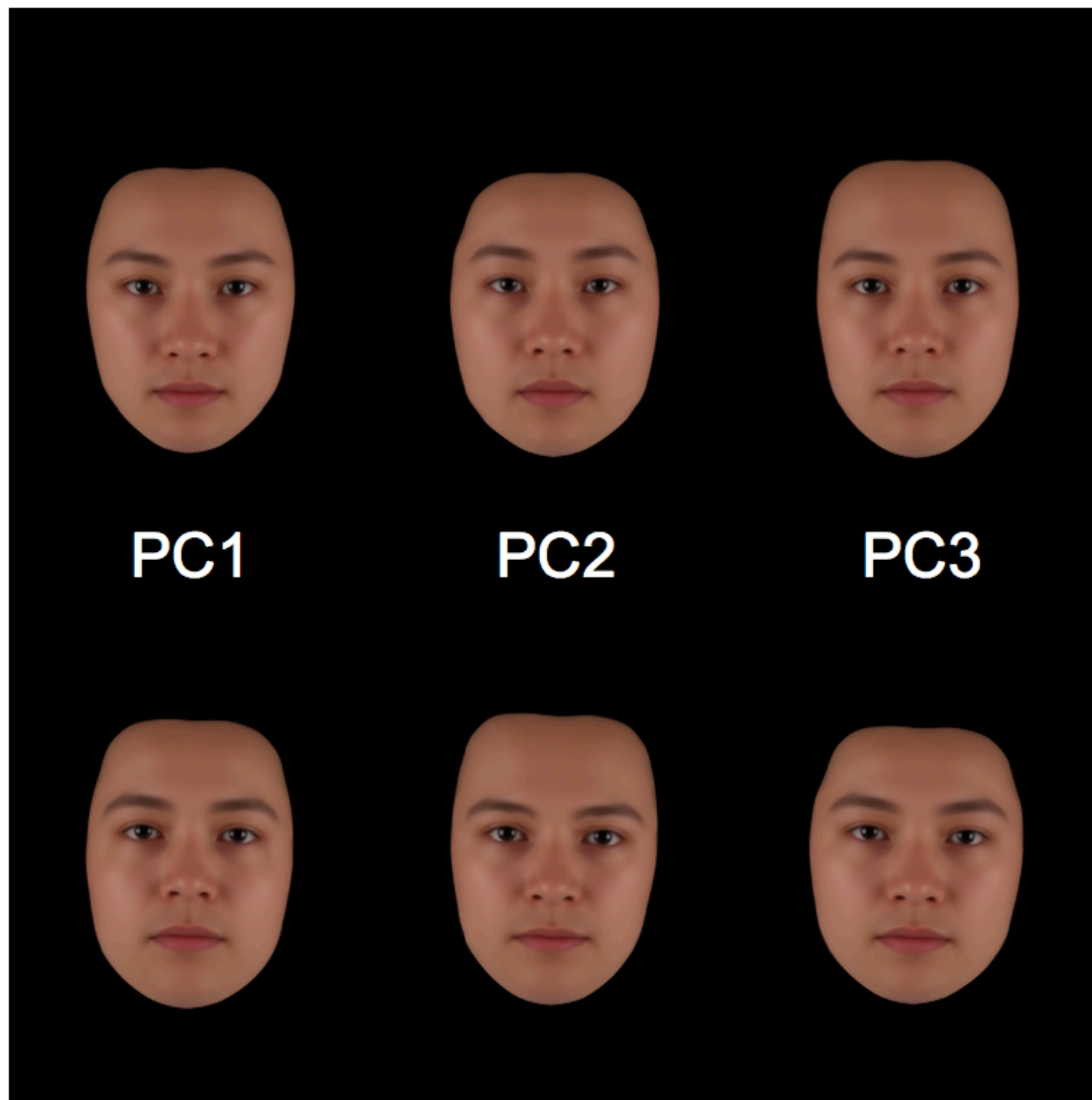


Figure 1. Visualization of the first three PCs. These three PCs explained 48% of the variance in 2D face shape. Components are applied to the average face from the image set for visualizations. Each PC is visualized at +1.5SD (top) and -1.5SD (bottom). PC1, PC2, and PC3 appear to correspond primarily to head tilt and sexual dimorphism, face ethnicity, and elongation, respectively.

Statistical analyses

We had preregistered our analysis plan prior to data collection (<https://osf.io/7wy3t/>). However, the reviewers suggested that our preregistered models could be prone to overfitting. To address this concern, we have altered our analyses. The main difference between the analyses reported here and those outlined in our preregistration is to focus on the three

PCs that explained the most variance in 2D face shape (new analyses), rather than all PCs selected using the broken stick method (preregistered analyses). We report the analyses requested by the reviewers in the main manuscript (below), report our preregistered analyses in full on the OSF (<https://osf.io/7wy3t/>), and include a section at the end of our Results section describing the differences in the results from these two sets of analyses. Analyses were conducted using R version 3.4.2 (R Core Team, 2016), with lme4 version 1.1-13 (Bates et al., 2014) and lmerTest version 2.0-33 (Kuznetsova et al., 2013). Linear mixed models were required to take into account the non-independence of different groups' attractiveness ratings of the same stimuli. Separate linear mixed models were conducted for attractiveness ratings of male and female faces and for each combination of the three rater groups (Chinese China-resident raters, Chinese UK-resident raters, White UK raters) who rated those faces. In each model, predictors were the three PCs that explained the most variance in face shape, face ethnicity (effect coded: Chinese=0.5, White UK=-0.5), rater group (effect coded: see details for each model in the relevant Results subsections below), and all possible two- and three-way interactions. Full model specifications and full outputs are given in our Supplemental Materials. Data files and analysis scripts are publicly available on the Open Science Framework (<https://osf.io/7wy3t/>). For each model, we report Akaike Information Criterion (AIC) as a measure of model fit.

Results

Women's facial attractiveness

Model 1. The first model compared the effects of PCs on attractiveness ratings by Chinese UK-resident male raters (effect coded as 0.5) and White UK male raters (effect coded as -0.5). This analysis revealed significant interactions between rater group and both PC1 (standardized estimate = -0.20, $t = -3.45$, $p < .001$) and PC2 (standardized estimate = -0.23, $t = -3.55$, $p < .001$). Neither of these two-way interactions was qualified by a three-way interaction with face ethnicity (both absolute t s < 1.00 , both p s $> .32$). There were no significant effects involving PC3 (all absolute t s < 1.95 , all p s $> .05$).

Full results of this analysis are shown in Table 3. AIC for this model was 281.9.

Table 3. Full results of Model 1 (Chinese UK-resident male raters versus White UK male raters) for women's facial attractiveness.

	Standardized estimate	Standard error	t	p
PC1	0.13	0.08	1.67	.10
PC2	0.05	0.09	0.57	.57
PC3	-0.06	0.06	-1.00	.32
rater group	0.07	0.06	1.26	.21
face ethnicity	-0.11	0.16	-0.67	.51
PC1 x rater group	-0.20	0.06	-3.45	<.001
PC2 x rater group	-0.23	0.06	-3.55	<.001
PC3 x rater group	0.00	0.05	0.05	.96
PC1 x face ethnicity	0.13	0.16	0.83	.41
PC2 x face ethnicity	0.17	0.17	0.99	.32
PC3 x face ethnicity	-0.25	0.13	-1.95	.05
face ethnicity x rater group	-0.16	0.12	-1.36	.18
PC1 x face ethnicity x rater group	-0.03	0.11	-0.23	.82
PC2 x face ethnicity x rater group	-0.11	0.13	-0.90	.37
PC3 x face ethnicity x rater group	-0.05	0.09	-0.57	.57

Model 2. The second model compared the effects of PCs on attractiveness ratings by Chinese China-resident male raters (effect coded as 0.5) and White UK male raters (effect coded as -0.5). This analysis also revealed significant interactions between rater group and both PC1 (standardized estimate = -0.16, $t = -3.06$, $p < .01$) and PC2 (standardized estimate = -0.19, $t = -3.24$, p

< .01). Again, neither of these two-way interactions was qualified by a three-way interaction with face ethnicity (both absolute t s < 1.45, both p s > .15).

There were no significant effects involving rater group and PC3 (all absolute t s < 1.76, all p s > .08). Full results of this analysis are shown in Table 4. AIC for this model was 244.1.

Table 4. Full results of Model 2 (Chinese China-resident male raters versus White UK male raters) for women's facial attractiveness.

	Standardized estimate	Standard error	t	p
PC1	0.15	0.07	2.13	<0.05
PC2	0.07	0.08	0.86	.39
PC3	-0.04	0.06	-0.65	.51
rater group	0.03	0.06	0.57	.57
face ethnicity	-0.01	0.14	-0.06	.95
PC1 x rater group	-0.16	0.05	-3.06	<.01
PC2 x rater group	-0.19	0.06	-3.24	<.01
PC3 x rater group	0.06	0.04	1.29	.20
PC1 x face ethnicity	0.07	0.14	0.54	.59
PC2 x face ethnicity	0.14	0.16	0.93	.35
PC3 x face ethnicity	-0.20	0.11	-1.76	.08
face ethnicity x rater group	0.04	0.11	0.35	.73
PC1 x face ethnicity x rater group	-0.13	0.11	-1.28	.20
PC2 x face ethnicity x rater group	-0.17	0.12	-1.44	.15
PC3 x face ethnicity x rater group	0.05	0.09	0.53	.60

Thus, the results of our first and second models suggest that Chinese men (regardless of country of residence) used the information in PC1 (head tilt

and sexual dimorphism) and PC2 (face ethnicity) differently from White UK men when judging women's attractiveness.

Model 3. The third model compared the effects of PCs on attractiveness ratings by Chinese China-resident male raters (effect coded as 0.5) and Chinese UK-resident male raters (effect coded as -0.5). The two-way interactions between rater group and PC1 and PC2 that were significant in our first two models were not significant in this model (standardized estimate = -0.34, $t = 0.73$, $p = .47$; standardized estimate = 0.03, $t = 0.64$, $p = .52$). Full results of this analysis are shown in Table 5. AIC for this model was 241.6.

Table 5. Full results of Model 3 (Chinese UK-resident male raters versus Chinese China-resident) for women's facial attractiveness.

	Standardized estimate	Standard error	t	p
PC1	0.05	0.08	0.65	.52
PC2	-0.05	0.09	-0.54	.59
PC3	-0.04	0.06	-0.57	.57
rater group	-0.04	0.05	-0.87	.39
face ethnicity	-0.09	0.16	-0.56	.58
PC1 x rater group	0.03	0.05	0.73	.47
PC2 x rater group	0.03	0.05	0.64	.52
PC3 x rater group	0.05	0.04	1.39	.17
PC1 x face ethnicity	0.06	0.15	0.40	.69
PC2 x face ethnicity	0.09	0.17	0.51	.61
PC3 x face ethnicity	-0.23	0.12	-1.80	.07
face ethnicity x rater group	0.20	0.10	2.03	.05
PC1 x face ethnicity x rater group	-0.11	0.09	-1.15	.25
PC2 x face ethnicity x rater group	-0.06	0.11	-0.54	.59
PC3 x face ethnicity x rater group	0.10	0.08	1.27	.21

Men's facial attractiveness

We used the same three models to investigate women's judgments of men's facial attractiveness.

Model 1. The first model compared the effects of PCs on attractiveness ratings by Chinese UK-resident female raters (effect coded as 0.5) and White UK female raters (effect coded as -0.5). This analysis revealed no significant interactions involving rater group. Full results of this analysis are shown in Table 6. AIC for this model was 256.3.

Table 6. Full results of Model 1 (Chinese UK-resident female raters versus White UK female raters) for men's facial attractiveness.

	Standardized estimate	Standard error	t	p
PC1	0.13	0.05	2.72	<.001
PC2	-0.06	0.07	-0.85	.40
PC3	0.04	0.05	0.05	.78
rater group	-0.07	0.07	-1.07	.29
face ethnicity	0.32	0.17	1.92	.06
PC1 x rater group	-0.01	0.04	-0.20	.84
PC2 x rater group	-0.11	0.06	-1.82	.07
PC3 x rater group	0.06	0.04	1.69	.10
PC1 x face ethnicity	0.10	0.09	1.08	.28
PC2 x face ethnicity	-0.02	0.15	-0.16	.88
PC3 x face ethnicity	-0.28	0.09	-3.00	<.001
face ethnicity x rater group	0.00	0.13	0.03	.97
PC1 x face ethnicity x rater group	0.11	0.07	1.48	.14
PC2 x face ethnicity x rater group	-0.11	0.12	-0.95	.35
PC3 x face ethnicity x rater group	0.10	0.07	1.41	.16

Model 2. The second model compared the effects of PCs on attractiveness ratings by Chinese China-resident female raters (effect coded as 0.5) and White UK female raters (effect coded as -0.5). This model did not converge, so we ran a reduced model that excluded all three-way interactions. This model also did not converge, so we ran separate models for each PC. Each model initially included all main effects, two-way interactions, and three-way interactions involving the PC, rater group, and face ethnicity. These models converged for PC1 and PC3, but not PC2. Models for PC1 and PC3 showed no significant interactions between PC and rater group. A model for PC2 in

which the three-way interaction was removed converged and showed a significant interaction between PC2 and rater group (standardized estimate = -0.16, $t = -3.61$, $p = <.001$). This interaction suggested that the negative effect of PC2 on attractiveness was weaker in the White UK rater group. Full results of these analyses are shown in Tables 7, 8, and 9. Akaike information criterion for these models were all > 276 .

Table 7. Full results of models comparing Chinese China-resident female raters' and White UK female raters' ratings of men's facial attractiveness for PC1.

	Standardized estimate	Standard error	t	p
PC1	0.09	0.05	2.02	.05
rater group	0.00	0.05	-0.05	.96
face ethnicity	0.52	0.10	5.03	<.001
PC1 x rater group	-0.04	0.04	-0.88	.38
PC1 x face ethnicity	0.07	0.09	0.78	.44
face ethnicity x rater group	0.39	0.09	4.17	<.001
PC1 x face ethnicity x rater group	0.09	0.08	1.03	.31

Table 8. Full results of models comparing Chinese China-resident female raters' and White UK female raters' ratings of men's facial attractiveness for PC2.

	Standardized estimate	Standard error	t	p
PC2	-0.08	0.07	-1.08	.28
rater group	-0.04	0.05	-0.91	.37
face ethnicity	0.35	0.17	2.12	.04
PC2 x rater group	-0.16	0.04	-3.61	<.001
PC2 x face ethnicity	0.01	0.15	0.06	.96

Table 9. Full results of models comparing Chinese China-resident female raters' and White UK female raters' ratings of men's facial attractiveness for PC3.

	Standardized estimate	Standard error	t	p
PC3	0.04	0.05	0.97	.33
rater group	0.00	0.04	-0.02	.98
face ethnicity	0.45	0.10	4.53	<.001
PC3 x rater group	0.07	0.04	1.61	.11
PC3 x face ethnicity	-0.27	0.09	-2.94	<.001
face ethnicity x rater group	0.36	0.09	3.99	<.001
PC3 x face ethnicity x rater group	0.05	0.08	0.60	.55

Thus, the results of our first and second models suggest that there was little evidence that Chinese and White UK women differed in how they used face-information.

Model 3. The third model compared the effects of PCs on attractiveness ratings by Chinese China-resident female raters (effect coded as 0.5) and Chinese UK-resident female raters (effect coded as -0.5). This analysis revealed no significant interactions involving rater group. Full results of this analysis are shown in Table 10. AIC for this model was 204.2.

Table 10. Full results of Model 1 (Chinese UK-resident female raters versus Chinese China-resident female raters) for men's facial attractiveness.

	Standardized estimate	Standard error	t	p
PC1	0.10	0.05	2.26	.03
PC2	-0.10	0.07	-1.43	.16
PC3	0.07	0.04	1.53	.13
rater group	0.03	0.05	0.59	.56
face ethnicity	0.44	0.16	2.74	<.001
PC1 x rater group	-0.04	0.03	-1.50	.14
PC2 x rater group	0.03	0.05	0.61	.54
PC3 x rater group	0.00	0.03	0.07	.94
PC1 x face ethnicity	0.14	0.09	1.52	.13
PC2 x face ethnicity	-0.03	0.14	-0.24	.81
PC3 x face ethnicity	-0.24	0.09	-2.77	.01
face ethnicity x rater group	0.23	0.11	2.18	.03
PC1 x face ethnicity x rater group	-0.04	0.06	-0.65	.52
PC2 x face ethnicity x rater group	0.09	0.09	0.94	.35
PC3 x face ethnicity x rater group	-0.04	0.06	-0.67	.50

Differences between results of the analyses described above and those of our preregistered analyses

The primary difference between the results of the analyses described above and those produced by our preregistered analyses occurred for women's judgments of men's facial attractiveness. Differences in how White UK and Chinese women used PCs 2 and 3 that were significant in our preregistered analyses (see Supplemental Materials) were not significant in the analyses requested by the reviewers (i.e., the analyses described above).

Color PCs

We also conducted analyses of color PCs. Results of these analyses are given at <https://osf.io/7wy3t/>.

Discussion

The current study used a data-driven method (principal component analysis) to compare the face- information that Chinese and White UK participants use to make attractiveness judgments. Our analyses of men's ratings of women's facial attractiveness (both those suggested by the reviewers and those in our preregistered analysis plan) suggested that White UK men find both downward-tilted, more feminine female faces and female faces with Chinese face shapes more attractive than Chinese men do. Importantly, these effects were independent of the effects of stimulus ethnicity on attractiveness judgments, indicating they cannot simply be due to own-race biases in face processing.

By contrast with our results for men's ratings of women's facial attractiveness, evidence for cultural differences in how women used male face information was mixed. On one hand, the analyses requested by the reviewers showed little evidence for cultural differences in women's face preferences. On the other hand, our preregistered analyses suggest that Chinese women find male faces with White UK and more elongated faces more attractive than White UK women do. On the basis of these mixed results, we tentatively suggest that ethnicity and elongation of faces could be a fruitful line of inquiry in studies examining possible cultural differences in White UK and Chinese women's face preferences. Indeed, our data suggest Chinese women born in China showed stronger preferences for White male faces than the other rater-groups did.

We characterized PC1 as reflecting information regarding head tilt and sexual dimorphism. Disentangling these two aspects of faces in two-dimensional face images is not straightforward, since altering head tilt affects face proportions and altering face proportions alters apparent head tilt (see, e.g., Hehman et al., 2013 and Schneider et al., 2012). Regardless, even if PC1 did primarily

reflect head tilt, rather than facial morphology, this would not be uninteresting. Several lines of research have demonstrated the importance of variable aspects of facial appearance for facial attractiveness (e.g., Main et al., 2010), with some researchers arguing they are, in fact, more important for attractiveness judgments than morphological cues (Jenkins et al., 2011).

Unexpectedly, an interaction between face ethnicity and PC3 was present across all models (although not significant in all models). This interaction suggested that preferences for narrow faces were stronger for judgments of White UK faces than for judgments of Chinese faces. This result demonstrates that the effects of facial characteristics can vary according to the ethnicity of the face presented, in addition to the ethnicity of the rater.

Many researchers have hypothesized that cultural differences in face preferences occur because of differences in recent visual diet (i.e., are, at least partly, a consequence of cultural differences in the types of faces people have recently been exposed to, Little et al., 2011; Scott et al., 2014). This hypothesis is consistent with experimental evidence that face preferences can be rapidly recalibrated by viewing faces whose appearance was manipulated in a consistent way (e.g., to increase masculinity or feature-spacing, Little et al., 2005; Rhodes et al., 2003). In our study, we saw no evidence that UK-resident Chinese and Chinese-resident Chinese participants differed in their use of face information. This suggests that differences between Chinese and White UK participants' face preferences are not due to differences in recent visual experience. Although our data do not straightforwardly support the visual diet explanation of cultural differences in face preferences, our data cannot speak to the possibility that visual diet early in life calibrates face preferences and that this calibration is relatively robust to changes in visual diet that occur in adulthood (i.e., there may be a 'critical period' during development in which visual diet affects face preferences).

A potentially important limitation of the current study is that the majority of faces in our sample scored below the midpoint of the scale. In other words,

our sample included few faces that were considered highly attractive. Although this is not unusual for studies using standardized face stimuli (see, e.g., Bronstad et al., 2008; Kościński, 2013; Torrance et al., 2014; Wang et al., 2016), it means that our results may not necessarily generalize to judgments of highly attractive faces.

In summary, we used a data-driven method to compare how Chinese and White UK raters use information when assessing facial attractiveness. White UK men found downward-tilted, more feminine female faces and female faces with Chinese faces more attractive than Chinese men did. Evidence for cultural differences in women's use of male face information was mixed, however. Nonetheless, our data-driven approach to comparing attractiveness judgments revealed cross-cultural differences in face preferences that were not apparent in studies using more traditional, theory-driven approaches, at least for men's judgments of women's facial attractiveness.

References

- Apicella, C. L., Little, A. C., & Marlowe, F. W. (2007). Facial averageness and attractiveness in an isolated population of hunter-gatherers. *Perception* 36, 1813–1820.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1.1-13.
- Bronstad, P. M., Langlois, J. H., & Russell, R. (2008). Computational models of facial attractiveness judgments. *Perception*, 37(1), 126-142.
- Hancock, P. J., Bruce, V., & Burton, A. M. (1998). Testing principal component representations for faces. In 4th Neural Computation and Psychology Workshop, London, 9–11 April 1997 (pp. 84-97). Springer, London.
- Holzleitner, I. J., Lee, A. J., Hahn, A., Kandrik, M., Bovet, J., Renoult, J. P., ... & Jones, B. C. (2018). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. PsyArxiv.
- Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, 49(4), 747-754.

- Holzleitner, I. J., Hunter, D. W., Tiddeman, B. P., Seck, A., Re, D. E., & Perrett, D. I. (2014). Men's facial masculinity: When (body) size matters. *Perception*, 43(11), 1191-1202.
- Jack, R. E., Crivelli, C., & Wheatley, T. (in press). Data-driven methods to diversify knowledge of human psychology. *Trends in Cognitive Sciences*.
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P.G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 7241-7244.
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74, 2204-2214.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313-323.
- Komori, M., Kawamura, S., & Ishihara, S. (2011). Multiple mechanisms in the perception of face gender: Effect of sex-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 626.
- Kościński, K. (2013). Perception of facial attractiveness from static and dynamic stimuli. *Perception*, 42(2), 163-175.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-33.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390.
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B*, 366, 1638-1659.
- Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 2283-2287.

- Little, A. C., Apicella, C. L., & Marlowe, F. W. (2007). Preferences for symmetry in human faces in two cultures: data from the UK and the Hadza, an isolated group of hunter-gatherers. *Proceedings of the Royal Society of London B*, 274, 3113-3117.
- Little, A. C., Jones, B. C., DeBruine, L. M. & Caldwell, C. C. (2011). Social learning and human mate preferences: a potential mechanism for generating and maintaining between-population diversity in attraction. *Philosophical Transactions of the Royal Society B*, 366, 366-375.
- Main, J. C., DeBruine, L. M., Little, A. C., & Jones, B. C. (2010). Interactions among the effects of head orientation, emotional expression, and physical attractiveness on face preferences. *Perception*, 39(1), 62-71.
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., & Castles, D. L., & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, 394, 884-887.
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199-226.
- Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception*, 30, 611-625.
- Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science*, 14, 558-566.
- Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*, 22, 1183-1190.
- Schneider, T. M., Hecht, H., & Carbon, C. C. (2012). Judging body weight from faces: The height—weight illusion. *Perception*, 41(1), 121-124.
- Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., ... & Honey, P. L. (2014). Human preferences for sexually dimorphic faces may be evolutionarily novel. *Proceedings of the National Academy of Sciences*, 111, 14388-14393.
- Scott, I. M., Pound, N., Stephen, I. D., Clark, A. P., & Penton-Voak, I. S. (2010). Does masculinity matter? The contribution of masculine face shape to male attractiveness in humans. *PLoS one*, 5(10), e13585.

- Thornhill, R. & Gangestad, S. W. (1999). Facial Attractiveness. *Trends in Cognitive Sciences*, 3, 452 – 460
- Torrance, J. S., Wincenciak, J., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). The relative contributions of facial shape and surface information to perceptions of attractiveness and dominance. *PloS One*, 9(10), e104415.
- Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). The motivational salience of faces is related to both their valence and dominance. *PloS one*, 11(8), e0161114.
- Wolffhechel, K., Hahn, A. C., Jarmer, H., Fisher, C. I., Jones, B. C., & DeBruine, L. M. (2015) Testing the utility of a data-driven approach for assessing BMI from face images. *PLoS ONE*, 10, e0140347.